

RECEIVED

MAR 28 2003

TECH CENTER



1600

ENTERED

## RAW SEQUENCE LISTING

DATE: 03/19/2003

PATENT APPLICATION: US/09/830,691A

TIME: 14:36:22

Input Set : A:\Seqlist.txt

Output Set: N:\CRF4\03192003\I830691A.raw

```
4 <110> APPLICANT: Choi, Eui-Sung
5      Rhee, Sang-Ki
6      Sohn, Jung-Hoon
7      Park, Soo-Dong
8      Lee, Yoon-Hyoung
9      Lee, Seung-Jae
10     Jang, Jae-Kweon
11     Choi, Seok-Keun
12     Son, Young-Rok
14 <120> TITLE OF INVENTION: VECTOR FOR THE TRANSFORMATION OF PHAFFIA
15     RHODOZYMA AND PROCESS OF TRANSFORMATION THEREBY
18 <130> FILE REFERENCE: 118.12-US-WO
20 <140> CURRENT APPLICATION NUMBER: 09/830,691A
21 <141> CURRENT FILING DATE: 2001-04-26
23 <150> PRIOR APPLICATION NUMBER: KR 1998/46547
24 <151> PRIOR FILING DATE: 1998-10-31
26 <150> PRIOR APPLICATION NUMBER: PCT/KR99/00265
27 <151> PRIOR FILING DATE: 1999-05-29
29 <160> NUMBER OF SEQ ID NOS: 20
31 <170> SOFTWARE: FastSEQ for Windows Version 4.0
33 <210> SEQ ID NO: 1
34 <211> LENGTH: 1223
35 <212> TYPE: DNA
36 <213> ORGANISM: Phaffia rhodozyma
38 <400> SEQUENCE: 1
39 atggtcaacg ttccaagac tcgacgtgag ttatagcaat ttcaacaact ctccagacga 60
40 caaatattcc agtgcacga aagagtttgt ggataaacgc gacagtttca agggaaagag 120
41 tcgatggaca gatttggaag acttagccgg tcaaggaact tggggatcac gtggcggagg 180
42 actcatcaga agaagtcggg atttgtttga tcatagtggg atcaagacaa actggaggat 240
43 atggctcgcc ttggaaggga atctccggcc tggattcgag gatccgaaag ttgtacgtat 300
44 ggaaaagctt acacggcttg gatttattat ctttcatagg aacctactgc aagggtaagg 360
45 cttgcaagaa gcacacgtaa gtcgcttatt ctctccactc tttcatggca tattgtcaac 420
46 gactggacaa cgcgtccgtt ttgaaacaag tgacttacct gtgaaatttg attctacacc 480
47 tgtatttagc cctcacaaag tacatatcac atcctccacc cccaccctgc ccaacttctt 540
48 cagttcatct tgcctcgggt ttccacattc cctgatgacc tccttgtagt ttctttgcga 600
49 acgtttgttt ctgtttctgt aggtgaccca gtacaagaag ggaaaggact ccatcttcgc 660
50 ccagggaag cgacgatacg accgaaagca gtccgggttac ggaggtcaga ccaagcccgt 720
51 tttccacaag aaggctaaga ccaccaagaa ggtcgtcctt cgattggcgg tatttttgtt 780
52 tattttgaat tctttttgtg tatgcagact tttgatgatt atgctcctct gtcgtttttt 840
53 ctcttcaaac agagtgtccc gtctgcagtt cgttcttcct tccaacaaaa acttcaacta 900
54 cagacatcat aaacagacat cttacttcgg tgttctctct ttttttccgc agagtacaag 960
55 atgcagatga ccctcaagcg atgcaagcac ttcgagcttg gaggagacaa gaagaccaag 1020
56 ggttcgtctt ttgtccatat attctctggt tcaattctta tgttcctaac gtacttggtt 1080
```

## RAW SEQUENCE LISTING

DATE: 03/19/2003

PATENT APPLICATION: US/09/830,691A

TIME: 14:36:22

Input Set : A:\Seqlist.txt

Output Set: N:\CRF4\03192003\I830691A.raw

57 cctttttggt tcggatgttg tttctatcgg tgggtgttttc ttttctttgg atgcattatc 1140  
 58 atttatcgtg ttggactgtt ttcctctgct cgtttctttc tcctctgtac ttgtgcttct 1200  
 59 caggagccgc catctctttc taa 1223

61 &lt;210&gt; SEQ ID NO: 2

62 &lt;211&gt; LENGTH: 350

63 &lt;212&gt; TYPE: DNA

64 &lt;213&gt; ORGANISM: Phaffia rhodozyma

66 &lt;220&gt; FEATURE:

67 &lt;221&gt; NAME/KEY: CDS

68 &lt;222&gt; LOCATION: (30)...(347)

70 &lt;400&gt; SEQUENCE: 2

71 cccttcaagt ctgctctcaa tcagtcaag atg gtc aac gtt ccc aag act cga 53  
 72 Met Val Asn Val Pro Lys Thr Arg  
 73 1 5  
 75 cga acc tac tgc aag ggt aag gct tgc aag aag cac acc cct cac aag 101  
 76 Arg Thr Tyr Cys Lys Gly Lys Ala Cys Lys Lys His Thr Pro His Lys  
 77 10 15 20  
 79 gtg acc cag tac aag aag gga aag gac tcc atc ttc gcc cag gga aag 149  
 80 Val Thr Gln Tyr Lys Lys Gly Lys Asp Ser Ile Phe Ala Gln Gly Lys  
 81 25 30 35 40  
 83 cga cga tac gac cga aag cag tcc ggt tac gga ggt cag acc aag ccc 197  
 84 Arg Arg Tyr Asp Arg Lys Gln Ser Gly Tyr Gly Gly Gln Thr Lys Pro  
 85 45 50 55  
 87 gtt ttc cac aag aag gct aag acc acc aag aag gtc gtc ctt cga ttg 245  
 88 Val Phe His Lys Lys Ala Lys Thr Thr Lys Lys Val Val Leu Arg Leu  
 89 60 65 70  
 91 gag tgc tcc gtc tgc aag tac aag atg cag atg acc ctc aag cga tgc 293  
 92 Glu Cys Ser Val Cys Lys Tyr Lys Met Gln Met Thr Leu Lys Arg Cys  
 93 75 80 85  
 95 aag cac ttc gag ctt gga gga gac aag aag acc aag gga gcc gcc atc 341  
 96 Lys His Phe Glu Leu Gly Gly Asp Lys Lys Thr Lys Gly Ala Ala Ile  
 97 90 95 100  
 99 tct ttc taa 350

100 Ser Phe

101 105

104 &lt;210&gt; SEQ ID NO: 3

105 &lt;211&gt; LENGTH: 106

106 &lt;212&gt; TYPE: PRT

107 &lt;213&gt; ORGANISM: Phaffia rhodozyma

109 &lt;400&gt; SEQUENCE: 3

110 Met Val Asn Val Pro Lys Thr Arg Arg Thr Tyr Cys Lys Gly Lys Ala  
 111 1 5 10 15  
 112 Cys Lys Lys His Thr Pro His Lys Val Thr Gln Tyr Lys Lys Gly Lys  
 113 20 25 30  
 114 Asp Ser Ile Phe Ala Gln Gly Lys Arg Arg Tyr Asp Arg Lys Gln Ser  
 115 35 40 45  
 116 Gly Tyr Gly Gly Gln Thr Lys Pro Val Phe His Lys Lys Ala Lys Thr  
 117 50 55 60  
 118 Thr Lys Lys Val Val Leu Arg Leu Glu Cys Ser Val Cys Lys Tyr Lys

## RAW SEQUENCE LISTING

DATE: 03/19/2003

PATENT APPLICATION: US/09/830,691A

TIME: 14:36:22

Input Set : A:\Seqlist.txt

Output Set: N:\CRF4\03192003\I830691A.raw

```

119 65          70          75          80
120 Met Gln Met Thr Leu Lys Arg Cys Lys His Phe Glu Leu Gly Gly Asp
121          85          90          95
122 Lys Lys Thr Lys Lys Gly Ala Ala Ile Ser Phe
123          100          105
126 <210> SEQ ID NO: 4
127 <211> LENGTH: 741
128 <212> TYPE: DNA
129 <213> ORGANISM: Phaffia rhodozyma
131 <220> FEATURE:
132 <221> NAME/KEY: misc_feature
133 <222> LOCATION: (0)...(0)
134 <223> OTHER INFORMATION: n=a, t, c, or g
136 <400> SEQUENCE: 4
137 ctcgagtgga cggtaggaat ggcattcgtg tcgttggtgc tcaactcgaa cccaagcagt 60
138 cgcttaccgc gggtagcctc cgggtgggcg cgatgatttg tgggtgtgat tccttcacct 120
139 tgggtagaac gacgcgcaac caatcattcg gagaaccgct ccgtttagac cgaccagtct 180
140 gattgatcaa catgccagca cgtcctccgg gacggagact ggcggggatc gtacctcatc 240
141 tggaatcgct ggctcaatgg tagtagtctt cagcatcggc catgagggca gtctaggtgg 300
W--> 142 gttcgccctgc cgaagactgt gtgagtggtgc tganaactaa ttgagtaccg ggggataagg 360
143 caaggcgtgt ntggttgccg gtggctgtga gcgagtttgc tgcaaagcga ttcaatgcac 420
144 cccggcttgg ccagcgcgct gcgtcacgaa acacactaaa cggttgacgc cataaagtaa 480
145 taacacactc aagtttggtg tcccgggtgg gcctctgtgc ctgcgtggga cccgacggga 540
146 gaggaanaac ttctgtggcc ctctcctctg tggatagtta cctggttgat cctgccagta 600
147 gtcatatgct tgtctcaaag attagccat gcattgtctaa gtataaacia attcatactg 660
148 tgaaactgcg aatggctcat taaatcagtt atagtttatt tgatggtacc ttgctacatg 720
149 gataactgtg gtaattctag a 741
151 <210> SEQ ID NO: 5
152 <211> LENGTH: 23
153 <212> TYPE: DNA
154 <213> ORGANISM: Artificial Sequence
156 <220> FEATURE:
157 <223> OTHER INFORMATION: CYH1, a PCR primer for the cloning of L41 genomic
158 DNA fragment
W--> 160 <221> NAME/KEY: misc_feature
161 <222> LOCATION: (0)...(0)
162 <223> OTHER INFORMATION: n=a, t, c, or g
W--> 164 <400> 5
W--> 165 cgcgtagtta aygtncnaa rac 23
167 <210> SEQ ID NO: 6
168 <211> LENGTH: 25
169 <212> TYPE: DNA
170 <213> ORGANISM: Artificial Sequence
172 <220> FEATURE:
173 <223> OTHER INFORMATION: CYH3, a PCR primer for the cloning of L41 genomic
174 DNA fragment
176 <400> SEQUENCE: 6
177 cccgggtytt ggcytтыtr tgraa 25
179 <210> SEQ ID NO: 7

```

## RAW SEQUENCE LISTING

DATE: 03/19/2003

PATENT APPLICATION: US/09/830,691A

TIME: 14:36:22

Input Set : A:\Seqlist.txt

Output Set: N:\CRF4\03192003\I830691A.raw

```

180 <211> LENGTH: 24
181 <212> TYPE: DNA
182 <213> ORGANISM: Artificial Sequence
184 <220> FEATURE:
185 <223> OTHER INFORMATION: 3' RACE primer
187 <400> SEQUENCE: 7
188 ggtcagacca agcaagtttt tcac 24
190 <210> SEQ ID NO: 8
191 <211> LENGTH: 24
192 <212> TYPE: DNA
193 <213> ORGANISM: Artificial Sequence
195 <220> FEATURE:
196 <223> OTHER INFORMATION: 5' RACE primer
198 <400> SEQUENCE: 8
199 gtgaaaaact tgcttggtct gacc 24
201 <210> SEQ ID NO: 9
202 <211> LENGTH: 24
203 <212> TYPE: DNA
204 <213> ORGANISM: Artificial Sequence
206 <220> FEATURE:
207 <223> OTHER INFORMATION: sense primer for the mutagenesis of L41 gene
209 <400> SEQUENCE: 9
210 ggtcagacca agcaagtttt tcac 24
212 <210> SEQ ID NO: 10
213 <211> LENGTH: 24
214 <212> TYPE: DNA
215 <213> ORGANISM: Artificial Sequence
217 <220> FEATURE:
218 <223> OTHER INFORMATION: antisense primer for the mutagenesis of L41 gene
220 <400> SEQUENCE: 10
221 gtgaaaaact tgcttggtct gacc 24
223 <210> SEQ ID NO: 11
224 <211> LENGTH: 20
225 <212> TYPE: DNA
226 <213> ORGANISM: Artificial Sequence
228 <220> FEATURE:
229 <223> OTHER INFORMATION: a PCR primer corresponding to 18S rDNA
231 <400> SEQUENCE: 11
232 tcctagtaag cgcaagtcac 20
234 <210> SEQ ID NO: 12
235 <211> LENGTH: 20
236 <212> TYPE: DNA
237 <213> ORGANISM: Artificial Sequence
239 <220> FEATURE:
240 <223> OTHER INFORMATION: a PCR primer corresponding to 18S rDNA
242 <400> SEQUENCE: 12
243 ttcggccaag gaaagaaact 20
245 <210> SEQ ID NO: 13
246 <211> LENGTH: 20

```

## RAW SEQUENCE LISTING

DATE: 03/19/2003

PATENT APPLICATION: US/09/830,691A

TIME: 14:36:22

Input Set : A:\Seqlist.txt

Output Set: N:\CRF4\03192003\I830691A.raw

```

247 <212> TYPE: DNA
248 <213> ORGANISM: Artificial Sequence
250 <220> FEATURE:
251 <223> OTHER INFORMATION: a PCR primer corresponding to 28S rDNA
253 <400> SEQUENCE: 13
254 aatcggatta tccggagcta                                20
256 <210> SEQ ID NO: 14
257 <211> LENGTH: 20
258 <212> TYPE: DNA
259 <213> ORGANISM: Artificial Sequence
261 <220> FEATURE:
262 <223> OTHER INFORMATION: a PCR primer corresponding to 28S rDNA
264 <400> SEQUENCE: 14
265 gctataacac atccggagat                                20
267 <210> SEQ ID NO: 15
268 <211> LENGTH: 2192
269 <212> TYPE: DNA
270 <213> ORGANISM: Phaffia rhodozyma
272 <400> SEQUENCE: 15
273 aagagctatt tgaatgacga ccacaagagt gacgatcata ttgagcatag tataccaaaag 60
274 gccaaagggc tgtgtggtgt tctatgagtg gccttgatta tgtgttacat aaataaactg 120
275 atctcaattt ttcaaatact tgccaacact ttcatatatt cacacaaaaa aaagtcagat 180
276 tggcccacaa agtcagatac acgctcgatc gtcgacgggt tcaagcactt tgtcaggcga 240
277 aagaaaaggcc acagcaccac ccttcaagtc tcgtctcaat caggttcgtc tagctttttg 300
278 tgtgcaagga tttaccgtct tgatggattt gttcgttgaa agagaggaaa gaacatgctg 360
279 aactgacgaa agtgtgaaca aaaaattgtg attttttcat tgtgtttcgc tgggtctcctt 420
280 gctggggttg gttggatcgg atttatcttc tgtgttggat ggaaaaccct gaatgttctt 480
281 ttcttggaac tcttctaaac tcgacaaaac gattcattcc tccgtactgc tctggttctg 540
282 cctttttgaa tcgcatcgat aaattcttcc ctcggaacgt tcgatcaatc tccgtcaaac 600
283 ttatcatcca aaaatctctt ctcgactgcc gccttgctcc ttttcttcgt tctttcctta 660
284 atccgctttc gactaccctc cttctcttca cactcatagt caagatggtc aacgttccca 720
285 agactcgacg tgagttatag caatttcaac aactctccag acgacaaata ttccagtga 780
286 tcgaaagagt ttgtggataa acgcgacagt ttcaaggga agagtcgatg gacagatttg 840
287 gaagacttag ccggtcaagg aacttgggga tcacgtggcg gaggactcat cagaagaagt 900
288 cgggatttgt ttgatcatag tgggatcaag acaaactgga ggatatggct cgccttgga 960
289 gggaaatctc gccctggatt cgaggatccg aaagttgtac gtatggaaaa gcttacacgg 1020
290 cttggattta ttatctttca taggaaccta ctgcaagggt aaggcttgca agaagcacac 1080
291 gtaagtcgct tatcctctcc actctttcat ggcataattg caacgactgg acaacgcgtc 1140
292 cgttttgaaa caagtgaact acctgtgaaa ttgattcta cacctgtatt tagccctcac 1200
293 aaggtacata tcacatctc ccaccccacc ctgcccact tcttcagttc atcttgctct 1260
294 cgttttccac attccctgat gacctccttg tatgttcttt gcgaacgttt gtttctgttt 1320
295 ctgtaggtga cccagtacaa gaagggaag gactccatct tcgcccaggg aaagcgacga 1380
296 tacgaccgaa agcagtcgag ttacggaggt cagaccaagc ccgtttttca caagaaggct 1440
297 aagaccacca agaaggtcgt ccttcgattg ggtacgtttt tgtttatttt gaattctttt 1500
298 tgtgtatgca gacttttgat gattatgctc ctctgtcgtt ttttctcttc aaacagagtg 1560
299 ctccgtctgc agttcgttct tcttccaac caaaacttca actacagaca tcataaacag 1620
300 acatcttact tcggtgttct ctcttttttt ccgcagagta caagatgcag atgacctca 1680
301 agcgatgcaa gcacttcgag cttggaggag acaagaagac caagggttcg tcttttgtcc 1740
302 atatattctc tggttcactt cttatgttcc taacgtactt gtttcctttt tggttcggat 1800

```

RAW SEQUENCE LISTING ERROR SUMMARY      DATE: 03/19/2003  
PATENT APPLICATION: US/09/830,691A      TIME: 14:36:23

Input Set : A:\Seqlist.txt  
Output Set: N:\CRF4\03192003\I830691A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:4; N Pos. 334,371

Seq#:5; N Pos. 15,18

## VERIFICATION SUMMARY

DATE: 03/19/2003

PATENT APPLICATION: US/09/830,691A

TIME: 14:36:23

Input Set : A:\Seqlist.txt

Output Set: N:\CRF4\03192003\I830691A.raw

L:142 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4 after pos.:300

M:341 Repeated in SeqNo=4

L:160 M:281 W: Numeric Fields not Ordered, &lt;221&gt; Sort in ascending order!

L:164 M:258 W: Mandatory Feature missing, &lt;220&gt; Tag not found for SEQ ID#:5

L:165 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5 after pos.:0